MAX-PLANCK-GESELLSCHAFT

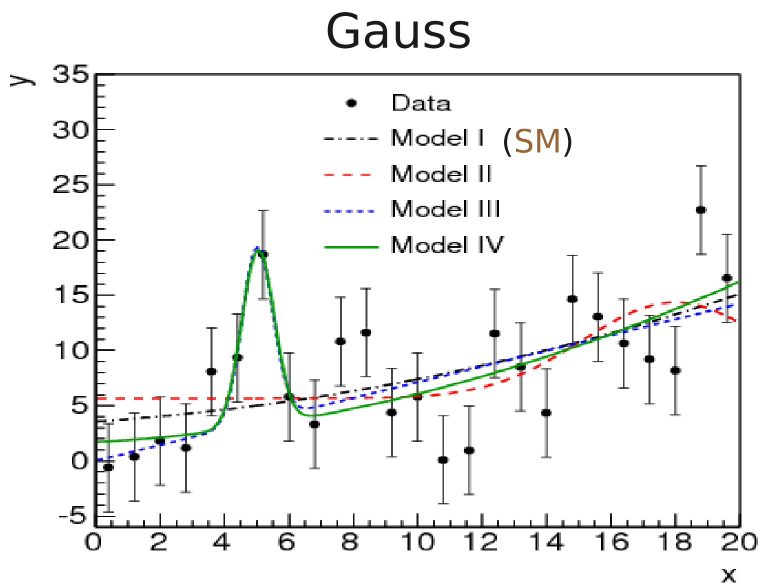
Max-Planck-Institut für Physik
(Werner-Heisenberg-Institut)

p-values for model evaluation

Frederik Beaujean (MPI für Physik)
Allen Caldwell (MPI für Physik)
Daniel Kollár (CERN)
Kevin Kröninger (Universität Göttingen)

DPG Frühjahrstagung 2011, Karlsruhe

Example problem

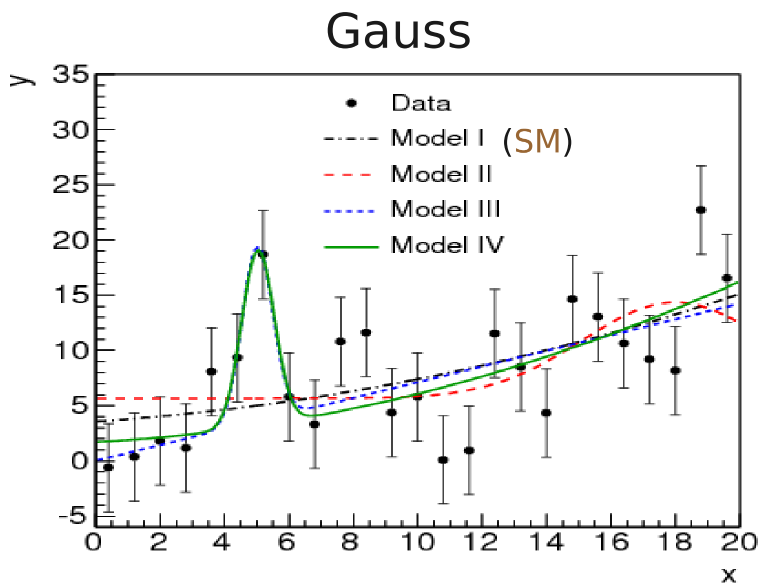


Suppose:

- *N* measurements (data points) with uncertainty
- Standard Model (SM) predicts quadratic background
- New physics (NP) predicts signal peak (more than one NP model)

Is Standard model enough to explain data?

Example problem



Fit function

$$y = f(x|\vec{\lambda}) = A + Bx + Cx^2 + \frac{D}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{\sigma^2}\right)$$

I : quadratic } SM
II : constant + Gaussian
III : linear + Gaussian
IV : quadratic + Gaussian } NP

Goodness of Fit: standard approach

Requirement:

- Assume a model *M* with parameters $\vec{\lambda}$

Test statistic:

- Any scalar function of data $T(D)$
- Interpret: large $T(D) =$ discrepancy between *M* and *D*

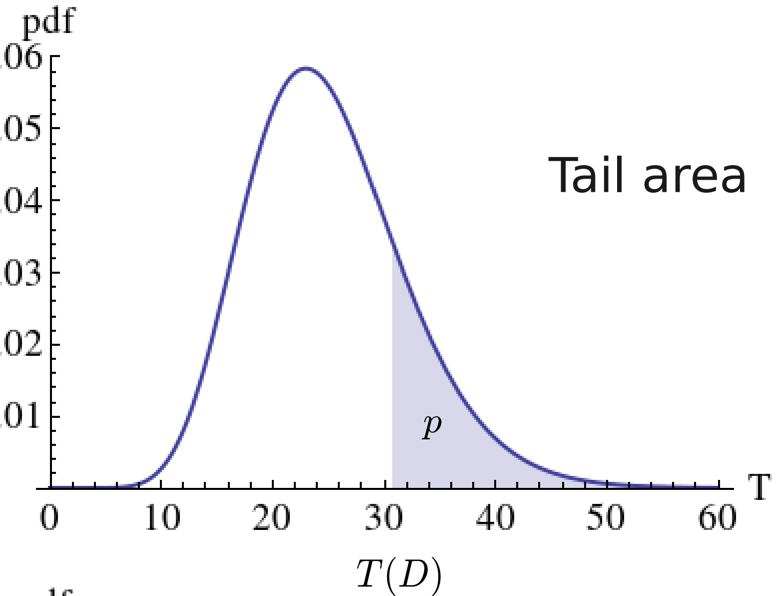
Example:

- Probability of the data $P(D|\vec{\lambda}, M) \propto \prod \exp\left\{ -\frac{\left(y_i - f(x_i|\vec{\lambda}, M)\right)^2}{2\sigma_i^2} \right\} = \exp\left\{ -\frac{\chi^2}{2} \right\}$
- Familiar choice $T(D) = \chi^2(D)$
- Extension: discrepancy variable $T(D|\vec{\lambda}, M)$. Fitting procedure important!

p-value

- Definition:

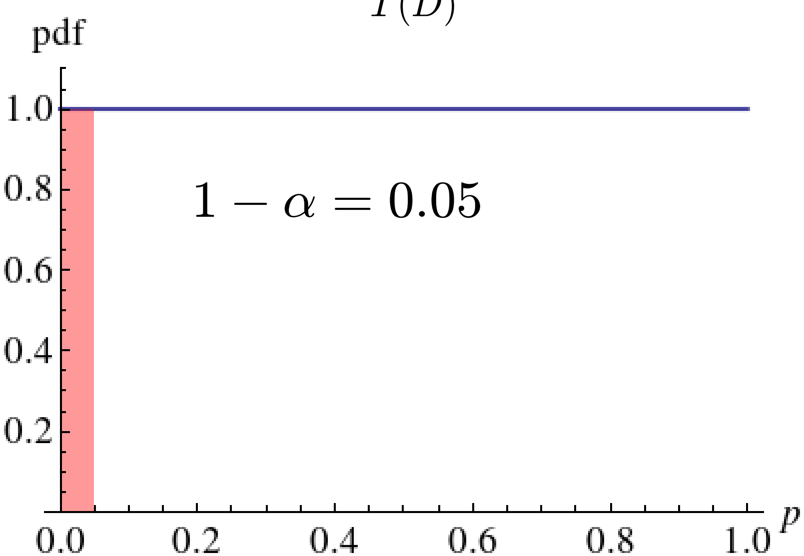
$$p \equiv P(T > T(D)|M)$$



- Assuming *M* and before data is taken: *p* uniform in [0,1]

- Confidence level α:

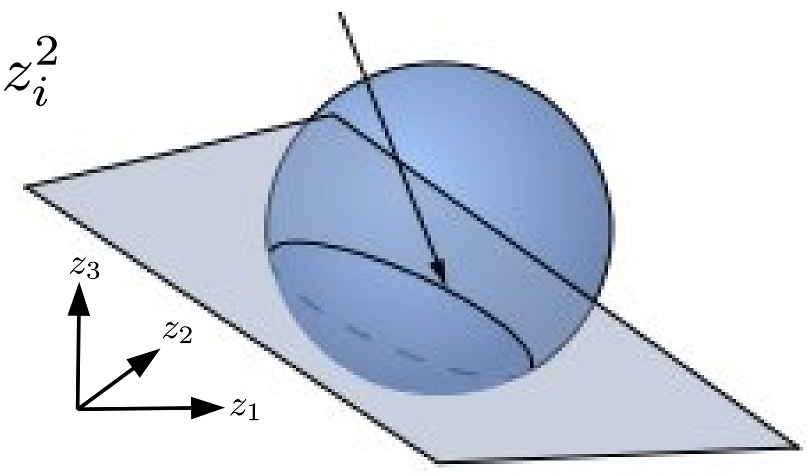
$$p < 1 - \alpha \Rightarrow \text{ reject model}$$



Gaussian linear regression

$$\chi^2(\vec{\lambda}, M) = \sum_{i=1}^{N} \frac{\left(f(x_i|\vec{\lambda}, M) - y_i\right)^2}{\sigma_i^2} = \sum_{i=1}^{N} z_i^2$$

Circle intersection



Least squares constraint,
find $\vec{\lambda}^*$ at global minimum:

$$\nabla\chi^2 \equiv \frac{\partial\chi^2}{\partial\lambda_j} = 0 \quad j = 1 \ldots k$$

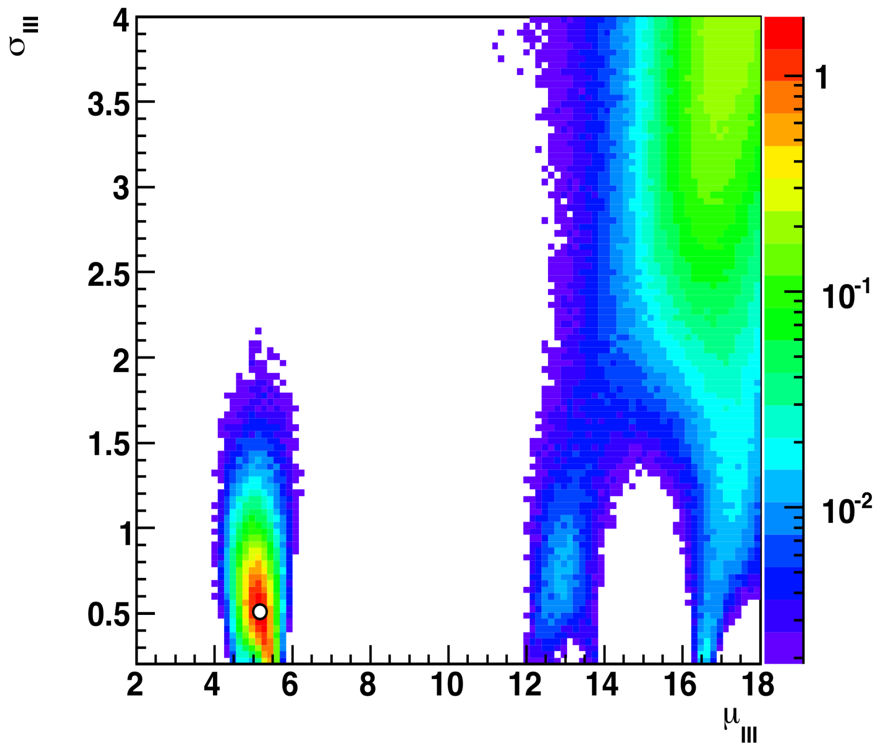
Predictions depend on parameters:

$$f(x_i|\vec{\lambda}^*, M) \text{ linear in } \vec{\lambda}^* \Rightarrow \nabla\chi^2 = 0 \text{ linear in } z_i \Rightarrow P\left(\chi^2|N - k \text{ DoF}\right)$$

Example: $f(x|\vec{\lambda}) = A + Bx + Cx^2 + \frac{D}{\sqrt{2\pi\ \sigma^2}} \exp\left(-\frac{(x-\mu)^2}{\sigma^2}\right)$ nonlinear!

In real life, usually $P(\chi^2|\vec{\lambda}^*, N, k) \neq P\left(\chi^2|N - k \text{ DoF}\right)$

Multimodality



Posterior of model III for particular data set and *small* range, flat priors

Two issues:
1) Find wrong mode within ranges
2) Global mode outside of ranges

- Physics motivates *small* parameter range U: e.g. C>0, σ>0.2 ..., but global mode possibly in *larger* range $V \supset U$

- Gradient based optimization (MINUIT/MIGRAD): need good starting point

- Clever user guess (difficult) or output from Monte Carlo sampler (preferred), e.g. Markov chain from Bayesian Analysis Toolkit BAT [mpp.mpg.de/bat/]

Comparison study

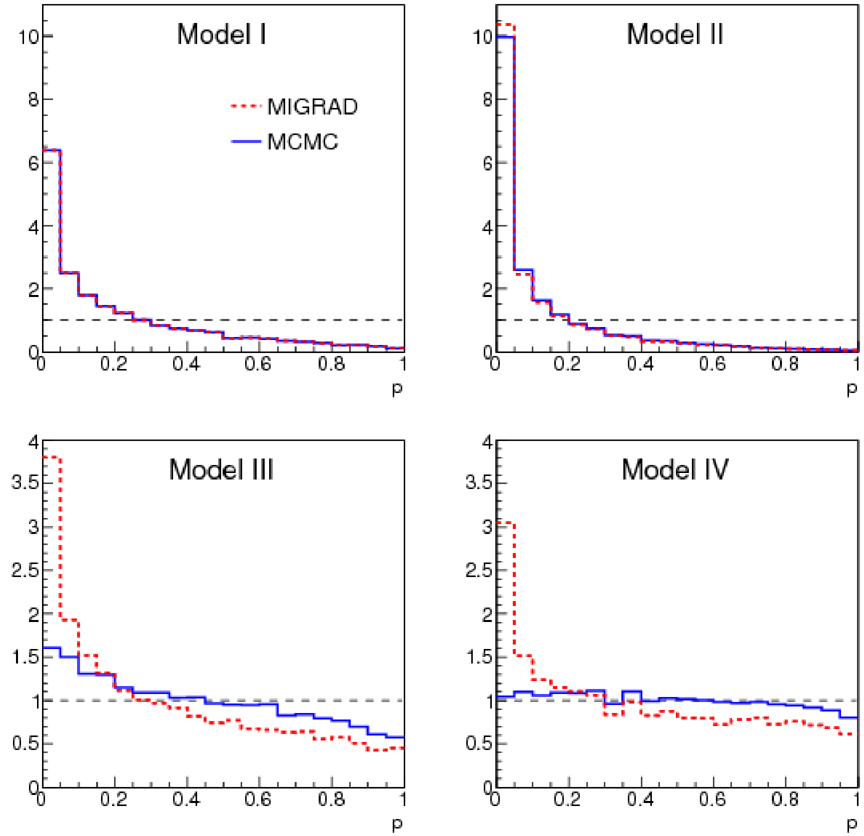
Goal: calculate p-value distribution for common statistics

- 10000 experiments
- Sample *N* data points from Model IV with fixed parameters
- Plot the distribution of the p-value for the statistics after fitting
- Restrict to Gaussian here

Beaujean, Caldwell, Kollár, Kröninger
Phys. Rev. D 83, 012004 (2011)

p-value distribution for χ^2 using (*N*-*k*) DoF

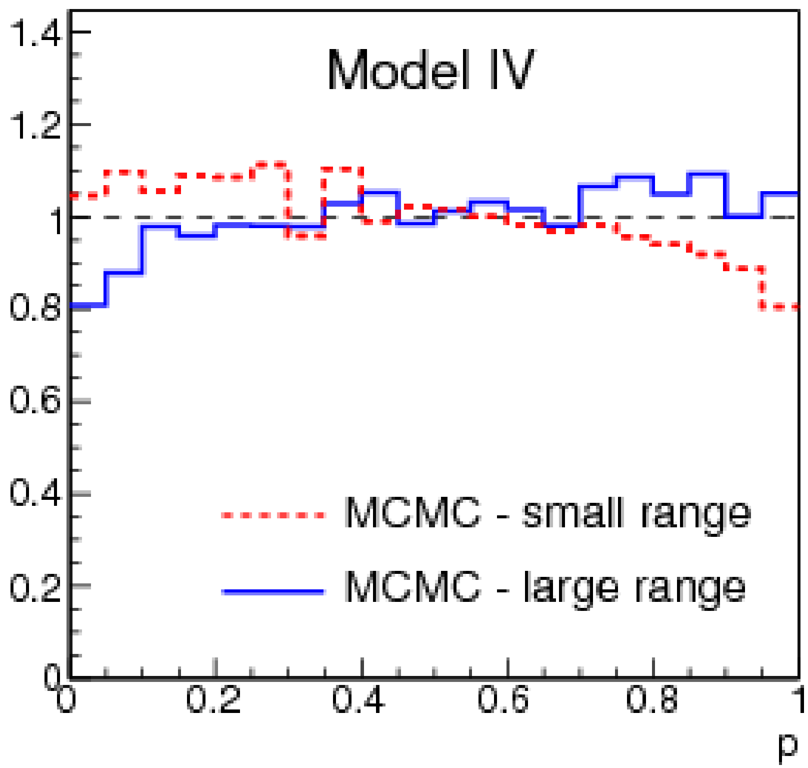
Small range U



- Models I and II ruled out.
- Models III and IV acceptable. Multimodality affects distribution.

MCMC+MINUIT much better than plain MINUIT

Small vs large range



- Small range: missing global minimum in some case, bias toward $p=0$

- True model, global minimum, but still distribution not flat.
  → Nonlinear fit function

Constraining parameter range = prior belief
Different prior → different p-value distribution

Conclusions

- *p*-values useful for goodness-of-fit
- Fitting can make big difference
- Beware: distributions usually approximate, keep uncertainty on *p*-value in mind

FINIS